

CLAIMS

What is claimed is:

- 5           1. A method of document transformation comprising:
- a) modeling a source XML document corresponding to a  
source schema as a source tree having a plurality of source  
nodes;
- b) modeling a target XML document corresponding to a  
10 target schema as a target tree having a plurality of target  
nodes; and
- c) generating a sequence of transformation operations  
that transforms said source tree to said target tree.
- 15           2. The method of document transformation as described  
in Claim 1, further comprising:
- d) converting said sequence of transformation  
operations into an Extensible Stylesheet Language for  
Transformations (XSLT) script.
- 20           3. The method of document transformation as described  
in Claim 1, wherein c) comprises:
- matching said plurality of source nodes to said  
plurality of target nodes.
- 25           4. The method of document transformation as described  
in Claim 1, wherein c) comprises:
- automatically generating said sequence of  
transformation operations.
- 30           5. The method of document transformation as described  
in Claim 1, further comprising:

d) for each source node in said source schema,  
selecting a plurality of candidate nodes in said target  
schema that are possible matches;

e) for each source node in said source schema,  
5 generating a plurality of node transformation sequences for  
transforming to each of said plurality of candidate nodes;  
and

f) for each source node in said source schema,  
selecting one of said plurality of node transformation  
10 sequences, a selected node transformation sequence, for said  
sequence of transformation operations that is associated  
with a least cost based on an information capacity cost  
criteria.

15 6. The method of document transformation as described  
in Claim 5, wherein f) further comprises:

in a match between a source node and a target node,  
selecting said selected node transformation sequence to  
achieve a high quality match, when an associated cost of  
20 data loss is less than a second cost of data loss when  
deleting information contained in said source node, in a  
first iteration of matching.

7. A method of document transformation as described  
25 in Claim 6, further comprising:

matching said source node to said target node having an  
identical label or synonymous label to achieve said high  
quality match.

30 8. The method of document transformation as described  
in Claim 5, wherein f) further comprises:

in a match between a source node and a target node,  
selecting said selected node transformation sequence when an

associated cost of data loss is less than a second cost of data loss when deleting source information contained in said source node and adding target information in said target node, in a second iteration of matching.

5

9. The method of document transformation as described in Claim 5, wherein f) further comprises:

selecting said selected node transformation sequence having the least associated cost of data loss.

10

10. A method of document transformation comprising:

a) modeling a source schema of XML and a target schema of XML as a tree structure creating a source tree and a target tree, said source tree having a plurality of source nodes, said target tree having a plurality of target nodes; and

15

b) generating a sequence of transformation operations that transforms said source XML document to said target XML document, wherein said plurality of source nodes of said source schema are matched and transformed to said plurality of target nodes in said target schema.

20

11. The method of document transformation as described in Claim 10, wherein b) comprises:

25

b1) for each source node in said source tree, selecting a plurality of candidate nodes in said target tree that are possible matches;

b2) for each source node in said source tree, generating a plurality of node transformation operations transforming to each of said plurality of candidate nodes; and

30

b3) for each source node in said source tree, selecting one of said plurality of node transformation

operations forming a selected node transformation operation having the least associated cost of information capacity.

12. The method of document transformation as  
5 described in Claim 11, further comprising:

combining said selected node transformation operation for each of said source nodes matched to a target node into a sequence of transformation operations that transforms said source schema to said target schema.

10

13. The method of document transformation as described in Claim 10, wherein said source schema is a source document type definition (DTD) and said target schema is a target DTD.

15

14. The method of document transformation as described in Claim 10, further comprising:

folding nodes in said source and target trees in a preprocessing phase to find one-to-one node matching.

20

15. The method of document transformation as described in Claim 10, further comprising:

merging nodes in said source and target trees in a preprocessing phase to find one-to-one node matching.

25

16. The method of document transformation as described in Claim 10, further comprising:

performing transformation operations only once at a node in said source tree and said target tree with the  
30 following exceptions:

- a) a relabel operation following an unfold operation;
- b) said unfold operation following said relabel operation;

c) said relabel operation performed between an attribute and an element following or followed by a deletion or an addition of a qmark quantifier node.

5           17. The method of document transformation as described in Claim 11, further comprising:

          converting said sequence of transformations operations into an Extensible Stylesheet Language for Transformations (XSLT) script.

10

          18. A computer system comprising:

          a processor; and

          a computer readable memory coupled to said processor and containing program instructions that, when executed,

15           implement a method of document transformation comprising:

          a) modeling a source XML document corresponding to a source schema as a source tree having a plurality of source nodes;

20           b) modeling a target XML document corresponding to a target schema as a target tree having a plurality of target nodes; and

          c) generating a sequence of transformation operations that transforms said source tree to said target tree.

25           19. The computer system as described in Claim 18, wherein said method further comprises:

          d) converting said sequence of transformation operations into an Extensible Stylesheet Language for Transformations (XSLT) script.

30

          20. The computer system as described in Claim 18, wherein c) in said method comprises:

matching said plurality of source nodes to said plurality of target nodes.

21. The computer system as described in Claim 18,  
5 wherein c) in said method comprises:

automatically generating said sequence of transformation operations.

22. The computer system as described in Claim 18,  
10 wherein said method further comprises:

d) for each source node in said source schema, selecting a plurality of candidate nodes in said target schema that are possible matches;

e) for each source node in said source schema,  
15 generating a plurality of node transformation sequences for transforming to each of said plurality of candidate nodes; and

f) for each source node in said source schema, selecting one of said plurality of node transformation  
20 sequences, a selected node transformation sequence, for said sequence of transformation operations that is associated with a least cost based on an information capacity cost criteria.

23. The computer system as described in Claim 22,  
25 wherein f) in said method further comprises:

in a match between a source node and a target node, selecting said selected node transformation sequence to achieve a high quality match, when an associated cost of  
30 data loss is less than a second cost of data loss when deleting information contained in said source node, in a first iteration of matching.

24. A computer system as described in Claim 23,  
wherein said method further comprises:

matching said source node to said target node having an  
identical label or synonymous label to achieve said high  
5 quality match.

25. The computer system as described in Claim 22,  
wherein f) in said method further comprises:

in a match between a source node and a target node,  
10 selecting said selected node transformation sequence when an  
associated cost of data loss is less than a second cost of  
data loss when deleting source information contained in said  
source node and adding target information in said target  
node, in a second iteration of matching. .

15 26. The computer system as described in Claim 22,  
wherein f) in said method further comprises:

selecting said selected node transformation sequence  
having the least associated cost of data loss.  
20